国际数据期刊出版的案例分析及其特征

■欧阳峥峥　青秀玲　顾立平　王　辉　吴　鸣

收稿日期:2014-07-30
修回日期:2015-03-24

中国科学院文献情报中心,北京中关村北四环西路 33 号　100190

摘　要　**【目的】**通过梳理国际代表性数据期刊特点,更好地了解数据期刊发展的趋势和特征。**【方法】**选择 15 种数据期刊进行调研,对其发展现状、学科分布、出版方式、收录内容、引用机制、版权协议以及出版费用等方面进行了分析。**【结果】**基于 *Scientific Data*、*F1000 Research*、*GigaScience* 等代表性期刊的分析,成功的数据期刊特点包括科研人员的认知度、学术社群的认可度、知识产权保护以及管理透明化。**【结论】**中国数据期刊出版界在搭建数据知识库和数据期刊出版平台后,还需关注如何提升国际认知度和认可度;图书馆员除了为科研人员推荐这些开放获取的数据资源外,更重要的是协助他们鉴别良好的数据期刊,以保护他们的知识产权权益。

关键词　数据期刊;数据论文;数据出版;数据引用;版权协议
DOI:10.11946/cjstp.201407300728

伴随着数据密集型科研活动的蓬勃发展,有一种活动愈来愈受重视——支撑关键发现的数据被更多的人访问,从而促进了数据的进一步分析以及结果的验证[1-2]。科学数据不仅仅是被研究的对象,其作为科研活动的主要产品“first class products”已经得到认可[3-4]。在此环境下,数据期刊(data journal)也快速发展起来。尤其是近年来,除开放获取出版商 Biomed Central 推出了数据期刊 *BMC Research Notes* 外,许多大型传统出版社也相继推出了数据期刊。如施普林格集团于 2012 年推出的 *SpringerPlus*,是一本跨学科的综合性开放获取期刊,其出版对象为原创研究、典型案例研究和研究方法的描述性文档以及数据论文和大数据集;Wiley 集团于 2012 年推出了 *Geoscience Data Journal*,用于发表地球科学数据论文,并通过数字对象标识符(DOI)将数据论文与存放在数据知识库中的数据集关联起来;自然出版集团 2014 年出版的 *Scientific Data* 旨在帮助科研人员发布、发现和重用研究数据,该杂志中一个重要的文章类型是“数据描述”(Data Descriptor),用于描述有科学价值的数据集,并帮助作者出版有科学价值和可重用性的数据集。在国内关于数据出版的研究中,刘凤红等人对数据论文的产生背景、概念、权益管理等问题进行了探讨,并提到了两种典型的数据期刊[5]。本文则对现有的数据期刊进行了调研,遴选部分具有代表性的数据期刊,分析了数据期刊的学科属性、出版方式、收录内容、引用机制、版权协议、费用等特点,指出了数据期刊的优势和面临的挑战。

1　国际数据期刊发展现状

数据期刊大致可分为两种,一种是纯粹的数据期刊,其出版对象全部为数据论文,如 *Scientific Data*、*Earth System Science Data*、*Open Health Data* 等;另一种是综合性的数据期刊,出版数据论文的同时也出版综述、研究论文、会议报告等其他类型的文献,如 *SpringerPlus*, *Biodiversity Data Journal*, *GigaScience* 等。而数据论文(data paper)是指按照学术规范正式出版的,可被检索的元数据文件,用以描述单个或一组可在线访问的数据集[6]。数据论文的内容主要是对数据采集、获取、处理等过程和方法的描述,不涉及对数据和研究结果的分析、推论、发现以及假设论证,其目的是让科研群体更好地发现、获取、理解与复用数据,并再次进行科研创新[7-8]。本文参考 Dyrad 网站[9]的线索,核实和

基金项目:中国科学院文献情报中心青年人才项目“多源开放数据在科研领域的应用研究”(项目编号:青 1305);科学数据管理服务试点建设(院 1522)。

第一作者简介:欧阳峥峥(ORCID:0000-0001-5941-3561),学科馆员, E-mail:ouyzz@mail.las.ac.cn

总结国际上已经开展数据出版的期刊共 85 种(网站更新日期 2014－10－10),调研了 15 种目前正式出版的数据期刊,如表 1 所示。

所调研的数据期刊,均有代表性,例如:BioMed Central 出版社在 2008 年较早进行了探索性试验,出版了开放获取期刊 *BMC Research Notes* 等,而国际著名学术出版商 Springer、Nature、Wiley、Elsevier 等也随后纷纷推出了数据期刊。从数据期刊出现的时间可以看出,自 2012 年起数据期刊进入发展高峰期。在开放获取的数据期刊中,尤以 *SpringerPlus*、*F1000 Research*、*BMC Research Notes* 等出版了大量的数据论文。

表 1　国际重要数据期刊的基本信息

年度	期刊名称	ISSN	出版商	发文总量(篇)
2009	*Earth System Science Data*[10]	1866－3508	Copernicus Publications	108
1999	*Ecology-data paper*[11]	0012－9658	Ecological Society of America	107
2014	*Genomics Data*[12]	2213－5960	Elsevier	89
2013	*Dataset Papers in Science*[13]	2314－8497	Hindawi publishing	51
2014	*Scientific Data*[14]	2052－4436	Nature Publication Group	37
2008	*BMC Research Notes*[15]	1756－0500	BoiMed Central	35
2009	*International Journal of Robotics Research*[16]	1741－3176	SAGE publications	13
2012	*Journal of Open Archaeology Data*[17]	2049－1565	Ubiquity Press	13
2013	*Open Health Data*[18]	2054－7102	Ubiquity Press	12
2012	*GigaScience*[19]	2047－217X	BoiMed Central	11
2012	*F1000 Research*[20]	2046－1402	F1000 Research	10
2013	*Biodiversity Data Journal*[21]	1314－2828	Pensoft Publishers	9
2013	*Journal of Open Psychology Data*[22]	2050－9863	Ubiquity Press	7
2014	*Geoscience Data Journal*[23]	2049－6060	Wiley Online Library	6
2014	*SpringerPlus*[24]	2193－1801	Springer	1

注:“年度”指首次出版数据论文时间;数量指发表数据论文的数量,单位为篇,统计时间为 2014 年 11 月 2 日。

2　国际数据期刊特点分析

本文对调研的 15 种期刊从学科属性、出版方式、收录内容要求、引用机制、版权协议、费用等方面进行分析。

2.1　学科分布

生命科学领域(含医学)的数据期刊发展较早,期刊种类较多,比较著名的数据期刊共有 6 种,分别是 *Genomics Data*、*BMC Research Notes*、*F1000 Research*、*GigaScience*[13]、*Open Health Data* 和 *Biodiversity Data Journal*。属于地球科学的数据期刊有 *Earth System Science Data* 和 *Geoscience Data Journal* 等。其它学科类别的数据期刊还包括生态学、心理学、计算机科学以及考古学等。数据期刊也有包含多个学科的综合性期刊,比较著名的有三种,分别是 *Scientific Data*、*Dataset Papers in Science* 以及 *SpringerPlus* 等。刘峰等人对 databib 上注册的数据知识库进行了学科属性分析,发现生物学方面的数据知识库数量最多[25],这可能是生命科学领域数据期刊发展较快的原因之一。从出版数据论文的时间来看,生态学期刊 *Ecology* 早在 1999 年就开始出版数据论文。该刊的数据论文作为一种特有的论文形式(data paper 栏目),用以发表该领域的大数据集以及描述大数据集内容、背景、结构、质量的元数据信息。地球学科的数据期刊出版形式均为数据论文,其中 2009 年正式出版的 *Earth System Science Data* 发表数据论文的数量最多。

2.2　出版方式

当前国际重要数据期刊普遍采取开放获取的方式正式出版(期刊 *Ecology* 虽然不是开放获取期刊,但其所有数据论文均可在线获取),以便于科学数据最大程度的传播和复用。在笔者所调查的数据期刊中,均采用同行评审机制,以保证期刊内容的质量以及可用性。与传统的期刊不同,大部分数据期刊仅以在线的方式出版(E-Only),少数期刊同时发行了纸本和电子版。而在出版周期上,若干数据期刊打破了按年、季、月、周等固定周期出版的模式,作者的数据论文只要通过同行评审,即可随时在线发表,笔者整理如表 2 所示。

表 2　国际重要数据期刊的比较

期刊名称	正式出版	同行评审	开放获取	出版载体		出版周期		
				在线	纸质	随时	年	月
Biodiversity Data Journal	●	●	●	●		●		
Dataset Papers in Science	●	●	●	●		●		
F1000 Research	●	●	●	●		●		
GigaScience	●	●	●	●		●		
Journal of Open Archaeology Data	●	●	●	●		●		
Journal of Open Psychology Data	●	●	●	●		●		
Open Health Data	●	●	●	●		●		
Scientific Data	●	●	●	●		●		
SpringerPlus	●	●	●	●		●		
Earth System Science Data	●	●	●	●	●		●	
Genomics Data	●	●	●	●			●	
BMC Research Notes	●	●	●	●				●
Geoscience Data Journal	●	●	●	●				●
Ecology-datapaper	●	●	●	●	●			●
International Journal of Robotics Research	●	●	●	●	●			●

2.3　收录内容

从数据期刊收录的内容和范围看,以数据论文为主要形式的期刊,即使有少数其他类型的文章,其内容也与数据处理、数据服务等相关。如 *GigaScience* 主要发表数据文献和一些介绍云计算、大数据领域软件与工具的文章;而 *Geoscience Data Journal* 主要发表数据论文以及关于数据服务和数据出版最佳实践的文章。另一类以数据论文作为子栏目之一的期刊,其收录范围除了数据论文外,还包括综述、研究论文、临床实验文章、观察文章、观点文章、研究记录、案例报告等,最具代表性的如 *BMC Research Notes*、*F1000 Research*、*SpringerPlus* 等。

从数据来源看,大部分期刊中数据论文的来源既可以是对已发表论文中数据的说明描述,也可以是对新产生数据的说明描述。当数据集较大时,需要先到指定或者推荐的数据知识库上将数据集注册,并将注册后的链接、标识符、元数据等信息放到数据论文中。存放数据集的数据知识库也必须具有开放共享性质,以保证数据集随时能够被访问和获取。此外,有的期刊还开发了自己的数据知识库,如 *GigaScience* 期刊中的数据集存放在该出版社开发的 Giga DB 数据知识库[26]中。这些具体要求,笔者整理如表 3 所示。

表 3　数据期刊的收录内容的具体要求

收录内容	具体要求
期刊收录范围	以数据论文为主导 数据论文作为期刊子栏目之一
数据注册	要求在数据知识库注册 未要求
数据保存	期刊提供推荐数据知识库列表 期刊自带数据知识库 期刊未指定数据知识库
数据保存仓储的开放性	数据知识库必须开放 未要求数据知识库开放

2.4　引用规范

所有期刊都有明确的引用要求,并给出了引用示例。如 Nature 出版的 *Scientific Data*,其引用格式包含作者、题目、期刊、卷、文章号、DOI 号、年代,具体示例为:

Falkenberg, K. J. et al. Genome-wide functional genomic and transcriptomic analyses for genes regulating sensitivity to vorinostat. Sci. Data 1:140017 doi: 10.1038/sdata.2014.17 (2014)”[27]

笔者所调查的数据期刊均采用数字对象标识符(DOI)对其数据论文进行唯一标识。在这些数据期刊中,有 10 种数据期刊在其引用要求中明确列出需要将 DOI 标识符作为引用字段。而其他 5 种数据期刊则未做此要求,其中 *Earth System Science Data* 期刊仅将 DOI 作为 Persistent ID(永久标识)字段的选择之一;*BMC Research Notes*、*GigaScience*、*Dataset Papers in Science* 用期刊中内部编码(如:article number)对引用文献进行标识;而 *Dataset Papers in Science* 则未提及唯一标识符。

2.5 版权协议

除 *Ecology* 的数据论文未提及作者版权归属外，其余的期刊均提到作者保留对其数据论文的著作权合法权利。而从出版协议看，数据期刊的主流形式是采用“知识共享许可协议”(Creative Commons License，又称创作共用许可，简称 CC)[28]作为许可协议。采用知识共享许可协议的期刊中，*Scientific Data*、*F1000 Research*、*Biodiversity Data Journal* 等采用最新版的 CC 4.0[29]。而 *Scientific Data* 与 *Genomics Data* 等还可以由作者自行选择开放层度不同的许可方式，如署名共享(CC BY 4.0)、署名 - 非商业性使用共享(CC BY-NC 4.0)、署名 - 非商业性使用 - 以相同方式共享(CC BY-NC-SA 4.0)。*Biodiversity Data Journal* 除了采用 CC 以外，其出版的科学数据还需遵照开放数据共享署名许可(Open Data Commons Attribution License)[30]。早期的数据期刊有采用该出版社自己设置的许可协议，例如 *BMC Research Notes* 早先采用了 Biomed Central 出版条款，后来随着整个出版社并入 Springer 集团，并且在积极推进开放获取出版的战略影响下，改为 CC 协议。出现类似情况的还有 ACS 出版协议、SAGE 出版协议等。由此可见，国外数据期刊的授权许可是在摸索中前进，可供我国出版界借鉴参考。

2.6 出版费用

除了 *Dataset Papers in Science*、*GigaScience*、*Biodiversity Data Journal* 三种期刊以外，笔者所调查的数据期刊大多要求支付数据出版费(Data Publishing Charge，DPC)，类似论文处理费(APC)。传统大型出版社的数据期刊出版费用较高，如 Nature 的 *Scientific Data* 期刊，如果选择 CC 4.0 授权许可，则作者需支付人民币 6850 元；Wiley 的 *Geoscience Data Journal* 期刊需要作者支付 1500 美元；BioMed Central 的 *BMC Research Notes* 期刊需要作者支付 1255 美元。而一些新兴的出版社，如 Ubiquity Press 出版的四种数据期刊则只是象征性收取 25 英镑，且如果作者无项目资助，还可以申请费用免除。同样可以申请费用免除的还有 Elsevier 集团出版的 *Genomics Data* 期刊。值得注意的是，许多目前免费的期刊都说明了免费的期限，因此收取一定的出版费用应该是未来数据期刊的发展趋势。我国作者和科研人员在投稿时，往往以发表论文作为第一考虑，对其中的“知识产权转让”问题和“出版费用支付”问题可能不太理解或者难以理解，这方面有待图书馆员进行相关培训和推广宣传。

3 数据期刊实例分析

为求深入理解一些已经成功的数据期刊案例，笔者仔细分析了由科研人员所推荐曾经投稿的数据期刊。整理如下：

3.1 *Scientific Data*

该刊是 Nature 集团 2013 年推出的综合性开放获取期刊，该刊主要以数据论文为主导，数据论文发表在其重要的文章类型“数据描述”(Data Descriptor)栏目下，发表后的数据论文被收录到一些重要的索引服务中被传播和重用。该刊提出了六个关键原则，这些原则也很好地概括了当前科研环境对数据期刊发展的要求：

1）信用：通过一个可被引用的出版物，储备和分享研究数据。

2）重用：完整、组织化、标准化的描述，保证数据可被重用。

3）质量：严格以学术团体为基础的同行评审。

4）发现：找到与研究相关的数据集。

5）开放：在数据使用、重用和分布上促进和支持开放的科学原则，通过知识共享许可协议获取。

6）服务：数据描述进行内部管理和快速同行评审。

3.2 *F1000 Research*

F1000 Research 是 2012 年开始启动的一个综合性的在线开放获取期刊，也是当前出版界创新模式的典型代表。主要发表生命科学和医学领域的新进展、新发现、科学数据、工具软件等，出版物类型包括数据论文、综述、研究论文、案例报告、临床实践文章、软件工具、观点性文章等 14 种。其特点如下：

1）发表数据论文：发表数据或数据论文是 *F1000 Research* 的一个重要特色，数据可以与研究论文一起发表，也可以单独以数据论文的形式发表，数据集可保存在 Dryad、Figshare 等与之合作的数据知识库中。

2）发表速度快：没有固定的出版周期，投稿经

过简单的语言和格式审查后,立刻公开发布在该网站上,并标记为"等待同行评审(awaiting peer review)",同行评审通过后,当天或第二天即正式发表在该期刊上,并获得一个 DOI 标识符,同时可被 scopus、google scholar 等检索。

3)后同行评审模式:稿件先提交在网站上公开发布,然后再进行同行评审,同行评审的结果及修改意见、专家观点等实时发布在网络上,作者可以同专家进行询问、解释、辩论。

4)过程公开透明:稿件通过简单语言和格式审核发布后,任何人都可以查看并发表评论,同行评审的意见、与作者的交流讨论以及每次修改后的稿件等都公布在网站上。

3.3 *GigaScience*

是由华大基因和 *BoiMed Central* 联合创办,面向所有使用或产出海量数据的生物学研究者的开放阅读期刊。

1)文献与数据库结合:期刊具有与之相关联的 GIGA DB 数据库,用于存放 *GigaScience* 期刊提到的数据集和工具。研究者在查看数据论文的时候可以直接利用数据集以及工具进行研究的重现与论证。

2)数据标识:GIGA DB 数据库采用 DOI 标识符对期刊中的所有数据进行标识,使数据保存更加永久,实现可追踪、可检索、可链接、可引用。

3)免费发表:到目前为止,用户在该期刊发表以及在对应数据库里对数据注册都无需支付费用。

3.4 *Earth System Science Data*

Earth System Science Data 于 2009 年开始出版地球科学相关数据,是一个以出版数据论文为主导且强调数据原创性的数据期刊。稿件可先通过简单快速的同行评审发表在 Earth System Science Data discussion 论坛上(该论坛也注册有 ISSN 号,并能形成正式引用),面向所有研究者开放并接受评论。待稿件正式通过同行评审后,则发表在 *Earth System Science Data* 期刊上。

4 数据期刊优势及挑战

随着科研第四范式的到来,科研环境对数据开放共享的要求越来越高。近几年来,数据期刊的发展越来越快,各大传统出版社纷纷推出了自己的数据期刊。结合前文对其特点的分析,数据期刊在数据发布和出版上,具有一些显著特点:

首先,数据期刊以开放获取作为主流的出版趋势,同时要求存放数据集和相关工具的仓储、平台也具有开放性,以保证科研群体在查看数据的同时可直接获取数据和工具,这使得科学数据在最大范围内得以传播、共享和复用。

其次,数据期刊不拘泥于固定周期的出版方式以及新颖的后同行评审模式提高了知识的传播速度。数据提前公开发表既避免了其他研究者无谓地重复,也使得这些数据可以在别人的研究中得到验证或进行新的发现,这些特点使得科研效率大幅度提高。

此外,在新颖的后同行评审(Post peer review)模式下,评审意见随同数据论文一起发布,作者可以进行讨论与辩解,科研群体也可公开发表评论和观点,所有信息的公开透明使得研究结果更加客观与公证。

更重要的是,数据期刊将数据以正式出版物的形式呈现,使得数据更规范、易理解、可引用,数据标识符的采用以及引用机制的建立使得数据生产者的贡献得以体现,数据的知识产权归属更加明晰,从而让研究者愿意共享数据,促进了科学数据的共享。

如果把数据期刊放在"数据发布与出版"的框架下,理解数据期刊在"数据驱动科研(data driven research)"范式中的作用与角色,那么它们的发展优势明显。然而,数据期刊可能尚未到达发展成熟的阶段,体现优势的同时也面临着一些挑战。

4.1 数据期刊认知度

数据期刊还未在科研群体中被广泛认知,这可从调研中大部分数据期刊上的发文量得到佐证。在我国,数据期刊尚处于起步和摸索的阶段, 2014 年在北京怀柔举办的科学数据大会上[31],中科院地理科学与资源研究所提出开始发行数据期刊《地理学报增刊》。这是我国第一本数据期刊,介绍有重要共享价值的科研数据,同时配合《全球变化科学研究数据注册与出版系统》[32]的运行,将数据存储在数据知识库中,并在数据期刊上发表。

但是,考虑到学术社群通用语言上的劣势、考核评价上关注国外期刊的现实情况以及国内科研

人员对于本国数据期刊和数据知识库尚未熟悉等各种条件,若要在我国出版界推动数据期刊的发展,可以考虑结合图书馆员的力量,提高这些优秀办刊刊物的国内认知度,以期在未来能够争取国际认知度。

4.2 数据期刊认可度

当前科研环境还缺乏对数据期刊以及数据论文的价值评价体系。尽管有一些用于评价数据期刊或数据论文价值的新技术和新方法出现,如 Thomson Reuters 集团新推出的数据引用索引数据库(Data citation index)[33]以及近两年开始尝试的基于文章层面的影响力计算方法(Article Level Metrics)[34-35],但这些新技术新方法目前仍处于尝试阶段,并未被纳入到研究者成果评价体系当中。研究者的成果无法像期刊论文一样,用期刊影响因子或者专家的评分(F1000 Prime)[36]来考量和计算,导致其贡献力和影响力无法在其科研共同体内得到认可和体现,这将大大降低研究者共享和发表数据的积极性。目前,国外已就这个问题展开各项研究与试验,如数据级别计量[37]的产生和推进。

4.3 知识产权保护

数据提供者和数据期刊在科学数据开放共享上,往往通过发布某种许可协议规范来引导其他科研人员如何使用、利用和重用这些发布或出版的科学数据。许可协议的适用范围是在著作权法以及邻接权的框架下进行,不同的许可协议有其适用条件和历史局限,需要随着数字技术发展、科学社群需求、认知的变迁以及法律法规和机构政策的调整而作出改变。目前通行的 CC 协议也只是其中一种途径,人们对此已经开展了许多宣传推广工作[38-39]。

数据引用在一定程度上,承认科研人员在收集、过滤、清理、整理、提供甚至共享科学数据上对知识创造和科学进步上的贡献。但是现有的数据引用方式还存在一定的局限,为了激发科学数据共享价值的最大潜力,人们正在进行一系列的尝试,试图从数据级别计量等的工作上,解决只能通过 DOI 或者数据期刊才能进行数据出版和引用的局限。

在论及科学数据的知识产权保护时,人们容易将著作权限制与例外、授权许可、数据引用进行混淆,其实它们属于不同范畴。如何保护科研人员以及科研教育机构在国外数据期刊上发表数据的权益,使其明确哪些权利应该转让、让渡与转移以及哪些权利需要保留,也是今后需要重点考虑的问题。

4.4 运营和管理

期刊的运营和管理需要成本,传统期刊的运营模式是一种由读者付费的运营模式,但传统模式也在逐渐转向开放获取出版模式[40]。对数据期刊而言,多数在一开始就采取开放获取出版的运营机制。该出版机制将评审费、编辑加工费、出版费、行政管理费等成本计算成 DPC 费用,要求作者及其所属的科研教育机构支付。与开放获取期刊评价与遴选[41]类似,数据期刊出版费的合理与否,要看它的投入成本与计算是否合理,而这方面就要求数据期刊的管理透明化。数据期刊与所有开放获取期刊一样,也面临着对其运营和管理模式的优化调整。

5 结语

数据期刊作为一种新兴的出版形式,其良好的发展需要科研、出版、数据服务、决策等多方人员的共同努力。出版者、数据服务者需要在科研群体中扩大数据期刊的认知度,让更多的研究者知晓、理解数据期刊的实质和意义。

中国科学院大学已经开设了相应的科学数据管理课程,围绕科学数据生命周期中的数据管理、发布、共享、复用等一系列问题进行了讨论,让越来越多的研究者了解到数据期刊。这将有助于未来的科研人员能够充分利用和识别良好的数据期刊。

此外,开发新技术、新方法建立对数据期刊或数据论文的引用与评价机制,同时梳理与科学数据相关的一系列政策也很有必要。中国科学院文献情报中心已经就此开展一系列政策研究工作,从科研资助机构、科研教育机构、信息服务机构、数据期刊、数据知识库、学术期刊的数据政策等,有系统地梳理政策要素,并且提供政策编译材料在机构知识库(ir. las. ac. cn)上开放共享。

目前由于种种因素,包括国外数据期刊政策本身的不明确等,我国科研人员在认识“知识产权转让”和“出版费用支付”的问题上,可能存在一些混淆。中国科学院文献情报中心的科技信息政策中心已经开始对这些内容进行分析,即将协同学科馆

员制定有关的政策调查报告及最佳实践指南,以便科研人员更加清楚地了解。

数据出版者则需要从多种途径来优化运营管理模式,如通过外部赞助、广告植入、开展增值服务等方式来减少自身运营投入,在最大程度上,将成本支出与收入来源以及经营管理予以透明化,在提升质量管理和刊物声誉的同时,让收取 DPC 的理由取信于人。总之,只有通过各方的努力和实践,数据期刊的优势才能更好地体现,从而更好地推动科研的发展。

致谢:感谢中国科学院文献情报中心初景利教授对本文的指导和建议。

参考文献

[1] Hey T, Tansley S, Tolle K. TheFourthParadigm: Data-IntensiveScientificDiscovery[M/OL]. [2012 - 03 - 08]. http://research. microsoft. com/en-us/collaboration/fourthparadigm/.

[2] 潘教峰,张晓林,等译. 第四范式:数据密集型科学发现[M]. 北京: 科学出版社, 2012: xvii.

[3] 张晓林. 开放获取、开放知识、开放创新推动开放知识服务模式—3O 会聚与研究图书馆范式再转变[J]. 现代图书情报技术, 2013, 29(2):1 - 10.

[4] Hans Pfeiffenberger. Data are First Class Products of Science-Data Journals are one Way to Establish this[EB/OL]. [2015 - 03 - 08]. http://oa. helmholtz. de/fileadmin/user_upload/redakteur/Workshops/helmholtz_oa_ws_som_14_pfeiffenberger. pdf

[5] 刘凤红,崔金钟,韩芳桥,等. 数据论文:大数据时代新兴学术论文出版类型探讨[J]. 中国科技期刊研究, 2014, 25(12): 1451 - 1456.

[6] Paul N, Peter C. Data papers-peer reviewed publication of high quality data sets [J]. *International Journal of Robotics Research*, 2009, 28(5):587.

[7] Candela L, Castelli D, Manghi P, et al. Data Journals: A Survey[J]. *Journal of the Association for Information Science Science and Technology*, 2015, doi: 10. 1002/asi. 23358 doi: 10. 1002/asi. 23358.

[8] Chavan V, Penev L. The data paper: a mechanism to incentivize data publishing in biodiversity science[J]. *BMC bioinformatics*, 2011, 12(Suppl 15):S2.

[9] Dyrad[EB/OL]. [2014- 07- 16]. http://www. datadryad. org/pages/integratedJournals.

[10] Earth System Science Data[EB/OL]. [2014-07-16]. http://www. earth-system-science-data. net/.

[11] Ecology[EB/OL]. [2014-07-16]. http://www. esajournals. org/loi/ecol.

[12] Genomics Data [EB/OL]. [2014- 07- 16]. http://www. journals. elsevier. com/genomics-data/.

[13] Dataset Papers in Science [EB/OL]. [2014- 07- 16] http://www. hindawi. com/journals/dpis/.

[14] Scientific Data [EB/OL]. [2014- 07- 16]. http://www. nature. com/sdata/.

[15] BMC Research Notes[EB/OL]. [2014- 07- 16]. http://www. biomedcentral. com/bmcresnotes.

[16] International Journal of Robotics Research[EB/OL]. [2014-07-16]. http://ijr. sagepub. com/.

[17] Journal of Open Archaeology Data[EB/OL]. [2014-07-16]. http://openarchaeologydata. metajnl. com/.

[18] Open Health Data [EB/OL]. [2014- 07- 16]. http://openhealthdata. metajnl. com/.

[19] GigaScience [EB/OL]. [2014- 07- 16]. http://www. gigasciencejournal. com/.

[20] F1000 Research [EB/OL]. [2014- 07- 16]. http://f1000research. com/.

[21] Biodiversity Data Journal[EB/OL]. [2014-07-16]. http://biodiversitydatajournal. com/.

[22] Journal of Open Psychology Data[EB/OL]. [2014-07-16]. http://openpsychologydata. metajnl. com/.

[23] Geoscience Data Journal [EB/OL]. [2014- 07- 16]. http://onlinelibrary. wiley. com/journal/10. 1002/(ISSN)2049 - 6060.

[24] SpringerPlus [EB/OL]. [2014- 07- 16]. http://www. springerplus. com/.

[25] 刘峰, 张晓林,孔丽华. 科研数据知识库研究述评[J]. 现代图书情报技术, 2014, 30(2):25 - 31.

[26] GIGA DB[EB/OL]. [2014-07-16]. http://gigadb. org/.

[27] citation example [2014- 07- 16] http://www. nature. com/articles/sdata201417.

[28] Creative Commons [EB/OL]. [2014- 07- 16]. http://creativecommons. org/.

[29] Creative Commons 4. 0[EB/OL]. [2014-07-16]. https://creativecommons. org/licenses/by/4. 0/.

[30] Open Data Commons Attribution License[EB/OL]. [2014-07-16]. http://opendatacommons. org/licenses/by/1. 0/.

[31] 2014 科学数据大会——科研大数据与数据科学[EB/OL]. [2014-07-16]. http://dc2014. codata. cn/dct/page/1.

[32] 全球变化科学研究数据注册与出版系统[EB/OL]. [2015 - 3 - 8]. http://www. geodoi. ac. cn/WebCn/Default. aspx.

[33] Data Citation Index [EB/OL]. [2014- 07- 16]. http://thomsonreuters. com/data-citation-index/.

[34] Article Level Metrics [EB/OL]. [2014- 07- 16]. http://article-level-metrics. plos. org/alm-info/.

[35] Tananbaum G. Article-Level Metrics: A SPARC Primer[OL]. (2013 - 04 - 17). [2013 - 06 - 08]. http://sparc. arl. org/sites/default/files/sparc-alm-primer. pdf.

[36] F1000 Prime[EB/OL]. [2014-07-16]. http://f1000.com/prime.
[37] 顾立平. 数据级别计量——概念辨析与实践进展[J]. 中国图书馆学报, 2015, 41(2):56-71.
[38] Milloy C, et al. Guide to Creative Commons [EB/OL]. [2013-08-16] http://oapen-uk.jiscebooks.org/ccguide/.
[39] 顾立平 译. 创作共用(知识共享) - 人文与社会科学的专著作者[EB/OL]. [2013-08-16] http://ir.las.ac.cn/handle/12502/6260.
[40] 张晓林,李麟,刘细文,等. 开放获取学术信息资源:逼近"主流化"转折点[J]. 图书情报工作,2012,56(9):42-47.
[41] 顾立平,张晓林,初景利,等. 开放获取期刊的评价与遴选:质量水平、开放程度和服务能力[J]. 图书情报工作,2013,57(1):49-54.

作者贡献声明:

欧阳峥峥:研究设计,信息采集与整理,论文撰写;
青秀玲:研究设计,信息补充,论文修改;
顾立平:信息补充,论文修改;
王　辉,吴　鸣:信息补充。

An study on the case analysis and characteristics of international data journal publishing

Ouyang Zhengzheng, QING Xiuling, KU Liping, WANG Hui, WU Ming

National Science Library, CAS, 33 Beisihuan xilu, Beijing 100190, China

Abstract: [**Purpose**] For better understanding the trends and features of the scientific data publishing and data journal development, this study focuses on the characteristics of the international representative data journals. [Methodlogy] We select and survey 15 data journals to analyze their development, subjects, publishing approaches, included contents, copyright agreements and publishing fees. [**Findings**] Based on the survey of significant journals such as Scientific Data, F1000 Research and GigaScience, the characteristics of successful data journals include the researchers' awareness, the recognition in the scientific community, the protection of intellectual property rights, and the management transparency. [**Conclusions**] The Chinese data journal publishing should concern how to improve international awareness and recognition after building data repositories and data journal publishing platforms. The librarian can not only suggest researchers using those open access data resources, but also assistant them in identifying good data journals in order to protect their IP rights.

Keywords: Data journal; Data article; Data publishing; Data citation; Copyright agreement

(本文责编:田　宏)